Structural conservation in the major facilitator superfamily as revealed by comparative modeling

EYAL VARDY,[1] ISAIAH T. ARKIN,[1] KAY E. GOTTSCHALK,[2] H. RONALD KABACK,[3] AND SHIMON SCHULDINER[1]

[1]Alexander Silberman Institute of Life Sciences, Hebrew University of Jerusalem, Jerusalem, 91904 Israel
[2]Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel
[3]Howard Hughes Medical Institute, Departments of Physiology and Microbiology, Immunology, and Molecular Genetics, Molecular Biology Institute, University of California, Los Angeles, California 90095, USA

(RECEIVED January 28, 2004; FINAL REVISION January 28, 2004; ACCEPTED April 7, 2004)

Abstract

The structures of membrane transporters are still mostly unsolved. Only recently, the first two high-resolution structures of transporters of the major facilitator superfamily (MFS) were published. Despite the low sequence similarity of the two proteins involved, lactose permease and glycerol-3-phosphate transporter, the reported structures are highly similar. This leads to the hypothesis that all members of the MFS share a similar structure, regardless of their low sequence identity. To test this hypothesis, we generated models of two other members of the MFS, the Tn10-encoded metal-tetracycline/H^+ antiporter (TetAB) and the rat vesicular monoamine transporter (rVMAT2). The models are based on the two MFS structures and on experimental data. The models for both proteins are in good agreement with the data available and support the notion of a shared fold for all MFS proteins.

Keywords: membrane protein; ion-coupled transporters; drug resistance; NEM accessibility

Transporters are responsible for creating and maintaining the different composition of the cell interior relative to the outside. The same is true for the solute gradient across internal organelles such as mitochondria, synaptic vesicles, and lysosomes. Thus, transporters are essential to sustain life and adaptation to changes in the environment. Their malfunction results in disease, for example, cystic fibrosis and cystinuria, which may lead to death. Because of their critical function, they are targets of therapeutic intervention; for example, the antidepressant Prozac is a blocker of the brain serotonin transporter (Blakely 2001). In some cases, they are responsible for the prevention of drug therapy; for example, multidrug transporters are responsible for the difficulties encountered in cancer chemotherapy and resistance of microorganisms to antibiotics (Putman et al. 2000; Ambudkar et al. 2003; Paulsen 2003).

Reprint requests to: Shimon Schuldiner, Alexander Silberman Institute of Life Sciences, Hebrew University of Jerusalem, Jerusalem, 91904 Israel; e-mail: Shimon.Schuldiner@huji.ac.il; fax: 972-2-5634625.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.04657704.

The major facilitator superfamily (MFS) represents the largest group of ion-coupled transporters (Saier et al. 1999). They translocate substrates against their electrochemical gradient by coupling the movement of an ion or a second solute down its gradient (Kaback et al. 2001). Despite intense interest in these proteins and a large number of laboratories experimenting with several of them, structural information at the atomic level was nonexistent until very recently.

The first structural evidence approaching atomic resolution was supplied by the interpretation of the three-dimensional structure of the oxalate transporter, OxlT, at a resolution of 6.5 Å (Hirai et al. 2002, 2003). By combining the structural information for OxlT with biochemical and sequence information from other MFS proteins, Subramanian, Maloney, and collaborators (Hirai et al. 2003) proposed a structural model. In their model, 12 helices surround a central cavity and form a symmetrical structure formed by two, six-helical lobes.

High-resolution structures of two proteins from this superfamily recently became available: the lactose permease

(LacY; Abramson et al. 2003) and the glycerol-3-phosphate transporter (GlpT; Huang et al. 2003) from *Escherichia coli*. The sequence identity between GlpT and LacY is only 21% and their mechanism of action appears to be different because the former uses the downhill gradient of phosphate for an obligatory exchange with glycerol-Pi (antiporter), whereas the latter cotransports β-galactosides with protons (symporter). Despite these differences, both structures show a highly similar fold, which is, in general, similar to the overall low-resolution structure of OxlT, on the basis of electron microscopy (EM) images (Hirai et al. 2002). Cα-RMSD between the structurally conserved regions of the two proteins is only 3.7 Å. Structural differences are observed in loop length and conformation. In addition, some of the trans membrane domains (TMs) of LacY are more kinked than in GlpT (Abramson et al. 2003; Huang et al. 2003).

The findings suggest the intriguing possibility that the fold of these transporters constitutes a scaffold for all MFS transporters with 12 helices. Although the fold is conserved, the specific function is obtained by varying sets of amino acids at the substrate binding and translocation domains. As observed earlier for other protein families, the fold appears to be better conserved than the sequence (Aronson et al. 1994; Rupasinghe et al. 2003).

In this work, we test this contention by simple comparative modeling of two experimentally characterized MFS proteins using the coordinates of the two known structures, LacY and GlpT, as templates. Both of the modeled proteins belong to the MFS. TetAB, the Tn10-encoded tetracycline/H^+ antiporter, has been exhaustively studied, and information about N-ethyl maleimide (NEM) accessibility of each one of the residues has been documented (Kimura et al. 1996, 1998; Konishi et al. 1999; Kimura-Someya et al. 2000; Tamura et al. 2001). The vesicular monoamine transporter (VMAT) is a mammalian transporter responsible for accumulation of monoamines in synaptic vesicles and chromaffin granules (Liu et al. 1992). VMAT is weakly but distinctly related to the multidrug transporters in the MFS (Liu et al. 1992; Schuldiner et al. 1995; Saier and Paulsen 2001).

The generated models are in excellent agreement with the available experimental data supporting the contention that structure is highly conserved in this superfamily.

Results

Examining the modeling approach

Models based on a manually improved sequence alignment of a target protein with each of the available templates were generated (Fig. 1). Although the fold of GlpT and LacY is highly similar (Fig. 2A), the low homology prevents the generation of a reliable alignment. Therefore, a manual optimization was performed as suggested by Sali, Marti-Renom, and collaborators (Marti-Renom et al. 2000). Cycles of computerized modeling followed by manual improvement of the alignment were done until a model that satisfies data and constraints was generated.

In order to examine the differences between a model based solely on experimental data, an optimal comparative model, and a real structure, models for GlpT and LacY were constructed. Examination of the models had a double purpose: (1) to find out the deviation of an optimal model from the real structure and (2) to estimate the accuracy of an alignment generated on the basis of multiple sequence alignment, TM prediction, and experimental data. Models based solely on sequence alignment were not expected to yield good results because of the very low sequence homology. Manual optimization of the alignment circumvented this problem to some extent and enabled generation of an acceptable model for LacY, as described following:

In order to get a good reference model, a structural alignment of GlpT and LacY was used first. The use of structural alignment enabled the generation of two models: the LacY model based on the GlpT structure, and the GlpT model based on the LacY structure. Because structural alignment relates residues by minimizing the RMSD between the two aligned structures, it is appropriate for comparative modeling.

Table 1. *Comparison of different structures and models using the Combinatorial Extension (CE) method*

	GlpT LacY	GlpT $GlpT_{SA}$	LacY $LacY_{SA}$	LacY $LacY_G$	$TetAB_L$ $TetAB_G$	$rVMAT2_L$ $rVMAT2_G$
RMSD total	3.7Å	3.6Å	3.9Å	3.9Å	3.8Å	3.7Å
RMSD TM1-6	3.4Å	3.3Å	3.4Å	3.3Å	3.4Å	3.3Å
RMSD TM7-12	3.0Å	3.0Å	3.2Å	2.9Å	3.1Å	2.9Å
Identity	9%	93%	98%	60%	29%	30%

Cα-RMSD for structural superimposition was calculated for whole protein and for each of its lobes (TM1-6 and TM7-12). In structural alignment, residues that correspond spatially are aligned, meaning that identity of structures with the same sequence represent the percent of residues with the same spatial placement (with the deviation of the structural fit) in both of the examined structures. (SA) Model by structural alignment; (L and G) model built according to LacY and GlpT on the basis of manually optimized sequence alignment.

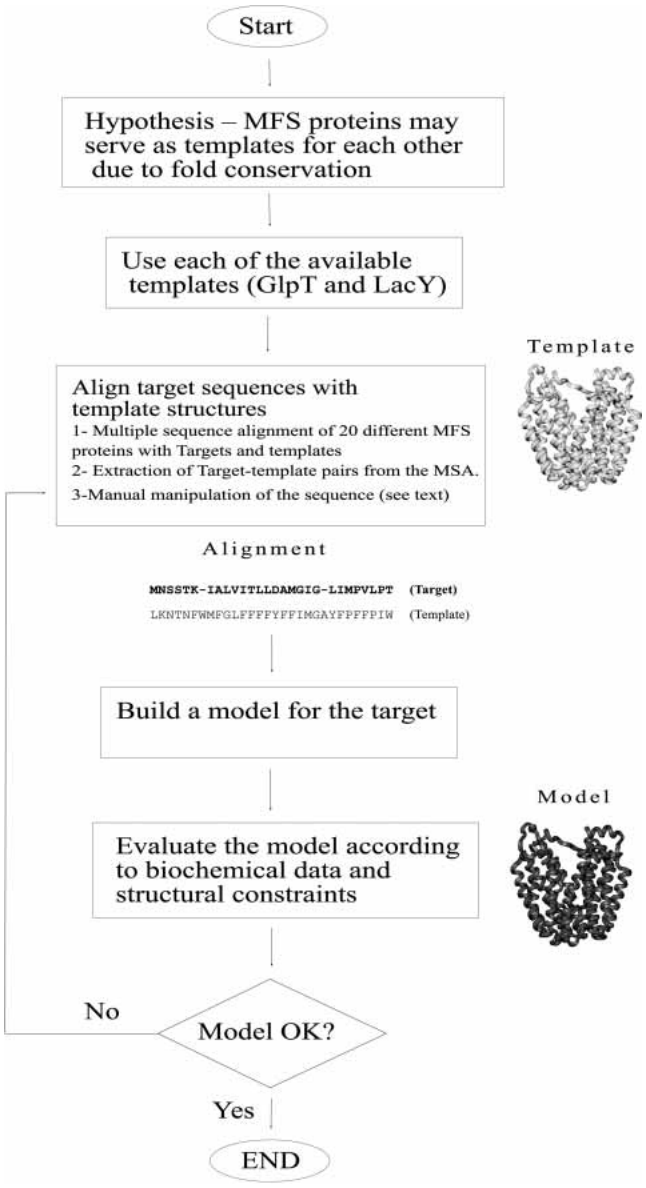

Figure 1. Flow chart of the modeling procedure. All sequence adjustments were done without insertions of gaps to TM regions (predicted or structural). Model examination was done by testing experimental data on the model. Cycles of manual optimization of the sequence alignment were continued until no further improvement could be achieved.

Each of these models was compared with the solved structure. Cα-RMSD for the structure versus the model was calculated for the two symmetrical halves of the protein (helices 1–6 and 7–12) and for the entire protein (Table 1). The position and orientation of membrane-embedded residues in the models were compared with the known structure (Fig. 2B,C). Superimposition of the models and structures produced, as expected, similar differences to superimposition of LacY and GlpT (~3.7 Å). The differences between the two halves of the proteins are smaller, especially of the C-terminal half (Table 1). That is mostly attributed to the fact that these proteins consist of two structurally similar lobes. The relative positioning of the lobes may change in different proteins and even at different conditions for the same protein. This mechanistic feature is made possible by the long central loop between TM6 and TM7 (L6–7) and has been suggested to be necessary for protein function (Weinglass and Kaback 2000).

Despite fold similarity of GlpT and LacY, there are differences between them that derive from different helix behavior and relative positioning. In LacY, TM4 is bent at two points and in GlpT this TM is almost completely straight, resulting in a 4 Å deviation of the two helices. Residues in TM5 in GlpT are shifted 3–8Å from the corresponding residues in the same TM in LacY, leading to a 5.8 Å Cα-RMSD of the two helices, but nevertheless the side chains in the TM regions of the models are mostly oriented as in the structure. These two examples represent the expected differences between a comparative model for MFS and its real structure when an optimal alignment is provided.

Modeling of LacY

In order to explore the capability of TM prediction combined with NEM accessibility data to generate an alignment

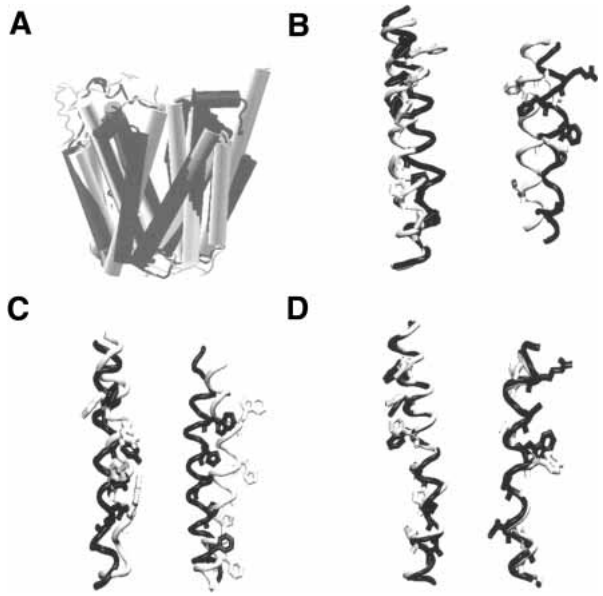

Figure 2. (*A*) Superimposition of GlpT and LacY structures. (*B*) Superimposition of LacY structure (bright) and model (according to structural alignment) by GlpT (dark). Shown are TM1 (*left*) and TM5 (*right*). (*C*) Superimposition of GlpT structure (bright) and model by LacY (dark). Shown are TM1 (*left*) and TM5 (*right*). Models in *B* and *C* are based on structural alignment. (*D*) Superimposition of two LacY models created according to structural alignment with GlpT (bright) or by manually optimized multiple sequence alignment (dark). Shown are TM1 (*left*) and TM5 (*right*). A phase shift between the modeled helices is observed in TM5.

good enough for comparative modeling, we built a model for LacY in three stages (Fig. 1): (1) Multiple sequence alignment was done for 20 MFS sequences from which the pairwise alignment of LacY (target) and GlpT (template) was extracted; (2) because of low homology, manual adjustment according to TM prediction was necessary; and (3) after modeling, the alignment was readjusted again, so that charged residues and NEM-accessible residues face the center of the protein (Venkatesan et al. 2000a,b,c; Kwaw et al. 2001). Six cycles of improvement were performed until a model that fits the constraints was achieved. The resulting model was then compared with the model built according to the structural alignment. The general fold of the two models is almost identical (the same template was used for both models) and most of the residues are positioned similarly in both models (Fig. 2D). Residues in TMs 4, 5, 6, and 8 are shifted by one position between the two models (a phase shift), but still all NEM accessibility data are compatible to both models. The phase shift is most probably caused by a faulty alignment, but because all NEM data are in agreement with the model, it might also reflect a structural substate with a rotated helix. The data available neither allow us to rectify the alignment nor to assign a functional importance.

Modeling of TetAB

The fold of GlpT and LacY is more conserved than their sequences. Based on the assumption that helix packing and fold of MFS is conserved and using alignments based on experimental data, models of TetAB with either LacY or GlpT as template were built. The preliminary alignments, derived from multiple sequence alignment, were manually improved, as was done for LacY modeling. The improved alignments were then used to generate models, which were in accordance with NEM accessibility data (Fig. 3A; Tamura et al. 2001) and the common helix packing pattern of MFS proteins with 12 TMs (Fig. 3B).

Structural alignment revealed that the two models are not identical even within the TM regions (Fig. 3D) and that most of the intervals between the sequences were three to four amino acids long, meaning that the differences between the models at these domains are of one helix turn (the helices are in phase), suggesting that helix orientation is similar. There are regions at the sequence alignment where a phase shift is observed inside the TM regions (± one or two residues). At those places (TM4, TM8, and TM11), structural differences between the helices of the two templates forced the phase shift and enabled compatibility of the experimental data. Thus, for example, the third turn in TM11 is of three residues in GlpT and four residues in LacY. The problem is overcome by a phase shift of the two models at that point.

As a consequence of our alignment procedure, all of the membrane-embedded charged residues in each of the models are facing the internal cavity. Because most of these residues are acidic, the cavity, according to the model, is mostly negative (Fig. 3C). The periphery of the models is mostly neutral, the cytoplasmic side is positively charged, and the periplasmic side is negatively charged. This is in accordance with the positive-inside rule (von Heijne and Gavel 1988).

Tilting and bending of the helices in both models is compatible with the NEM accessibility profile of the protein (Fig. 4C). The periplasmic half of TM4 and TM10 is blocked from NEM molecules by TM1–TM2 and TM7–TM8, respectively (Fig. 4A), whereas TM1 and TM7 are NEM inaccessible at the cytoplasmic half, where they are blocked by TM4–TM5 and TM10–TM11, respectively (Fig. 4B). The two different models are very similar (Cα-RMSD 3.8 Å) in the sense of fold and orientation of the residues in the helices, even though there are differences that derive from the structural differences of the two templates and alignment of TetAB with each of them.

Modeling of rVMAT2

Comparative modeling was used in the same manner as in the TetAB model to create models for another protein with less structural data. VMAT is a mammalian protein and its homology with other MFS proteins does not include L1–2, the vesicular loop between TM1 and TM2, which was not modeled. The sequence of rVMAT2 was aligned to GlpT and LacY on the basis of the same principles as TetAB. There are membrane-embedded charged residues in eight different TMs of rVMAT2, facilitating a correct orientation of those TMs in the model even in the absence of NEM accessibility data. The two generated models are as similar as the two starting points. Evaluation of the models was mainly by examining their feasibility in light of the position of charged residue in a functional context. All of the membrane-embedded charged residues face the water-filled cavity (Fig. 5A). Examination of the model reveals a potential ion pair that connects TM2 and TM11 (Fig. 5B). Experimental evidence for the existence of this ion pair between Asp 427 and Lys 139 was described (Merickel et al. 1997).

Vesicular acetylcholine transporter (VAChT) is a vesicular transporter closely related to VMAT (Schuldiner et al. 1995; Parsons 2000). Experimental evidence suggests high proximity of His 338 in TM8 of rVAChT and Asp 398 in TM10 (Kim et al. 2000). A multiple sequence alignment was calculated for VMAT and VAChT homologs (Fig. 5C). According to the alignment, His 338 is conserved in VAChT and is aligned against a conserved Tyr in VMATs. That Tyr (Y342) is positioned <3 Å from the Asp in TM10 in our model (Fig. 5D).

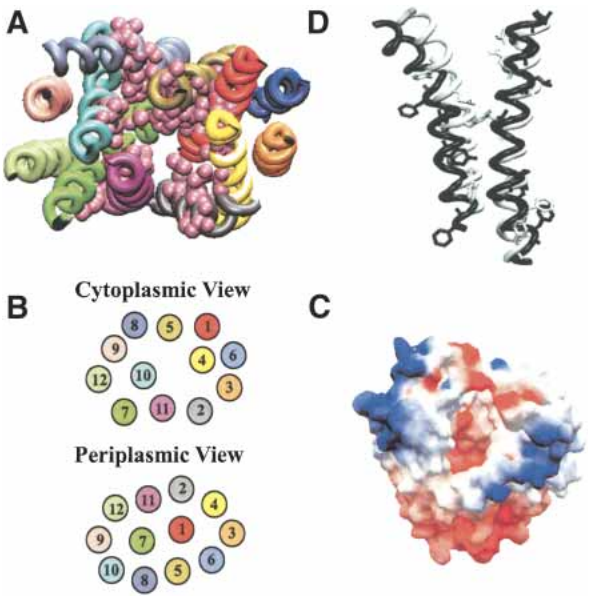

Figure 3. (*A*) Cytoplasmic view of NEM-accessible residues (mauve) in TetAB model (LacY as template). (*B*) Helix projection of TetAB model: cytoplasmic and periplasmic view. (*C*) Electrostatic surface representation of TetAB model. The cytoplasmic view reveals a negatively charged central cavity. (*D*) Superimposition of two TetAB models based on the two templates, LacY (white) and GlpT (black). Shown are TM2 (*left*) and TM11 (*right*). A phase shift of one to two residues is visible at the cytoplasmic half of TM11. That phase shift is corrected after the third turn, and the residues in the rest of the TM overlap. In TM2, a turn shift is observed. Each residue in TM2 of TetAB by GlpT is one helix turn closer to the periplasmic side of the membrane than the corresponding residue in TetAB by LacY.

Discussion

Examining the modeling approach

A sequence alignment based on structural similarity rather than sequential similarity served as input for comparative modeling. The structural alignment aligns spatially corresponding residues, thus generating an optimal alignment for comparative modeling. Models based on structural alignment of GlpT and LacY were compared with the solved structure. The models are, in general, similar to the known structures in the sense of protein folding, helix packing, and residue orientation. Nevertheless, certain differences exist. For instance, a shift of TM5 in the models relative to the structures can be observed. Even though that shift may be specific for modeling based on those two proteins, it represents the magnitude of the differences between comparative models of MFS and their real structures.

To get a more realistic estimate of correctness of our models, we generated a model for LacY, based on a manually optimized alignment (according to multiple sequence alignment, TM prediction, and experimental data) without considering structural alignment. Comparison of this model to a model derived from an ideal alignment showed that 70% of the manually optimized alignment is identical to the structural alignment (Table 1). The remaining 30% is located mainly in loops but also derived from a phase shift between TMs 4, 5, 6, and 8 of the two models. The data fit both models and the residues are positioned similarly despite the phase shift. From the comparison of the first "ideal" model and the second model, it can be deduced that the manual optimization of a sequence alignment of MFS proteins may generate an alignment very close to optimum. Models based on such an alignment may serve as insightful tools for biochemists studying MFS proteins.

TetAB model

Once the alignments were optimized, models were built for TetAB according to GlpT and LacY. The two models are mostly in agreement with NEM accessibility data (Figs. 3A, 4A–C). Residues 20, 24, 136, and 311 in the TetAB model

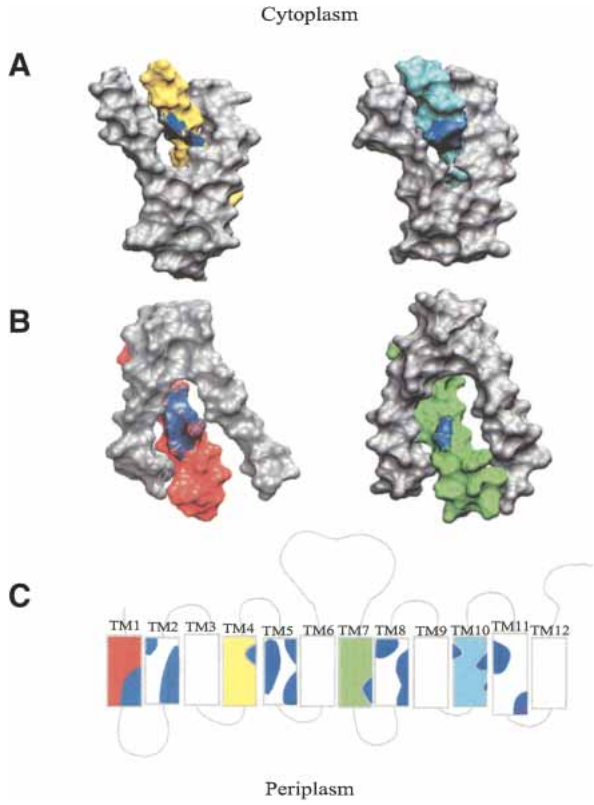

Figure 4. Blue areas represent NEM-accessible residues. (*A*) NEM accessibility to periplasmic half of TM4 (yellow) is prevented by TM1 and TM2. For TM10 (cyan), accessibility is prevented by TM7 and TM8. (*B*) NEM accessibility to cytoplasmic half of TM1 (red) is prevented by TM4 and TM5. For TM7 (green), accessibility is prevented by TM10 and TM11. (*C*) Two-dimensional representation of NEM accessibility profile in TetAB (Tamura et al. 2001).

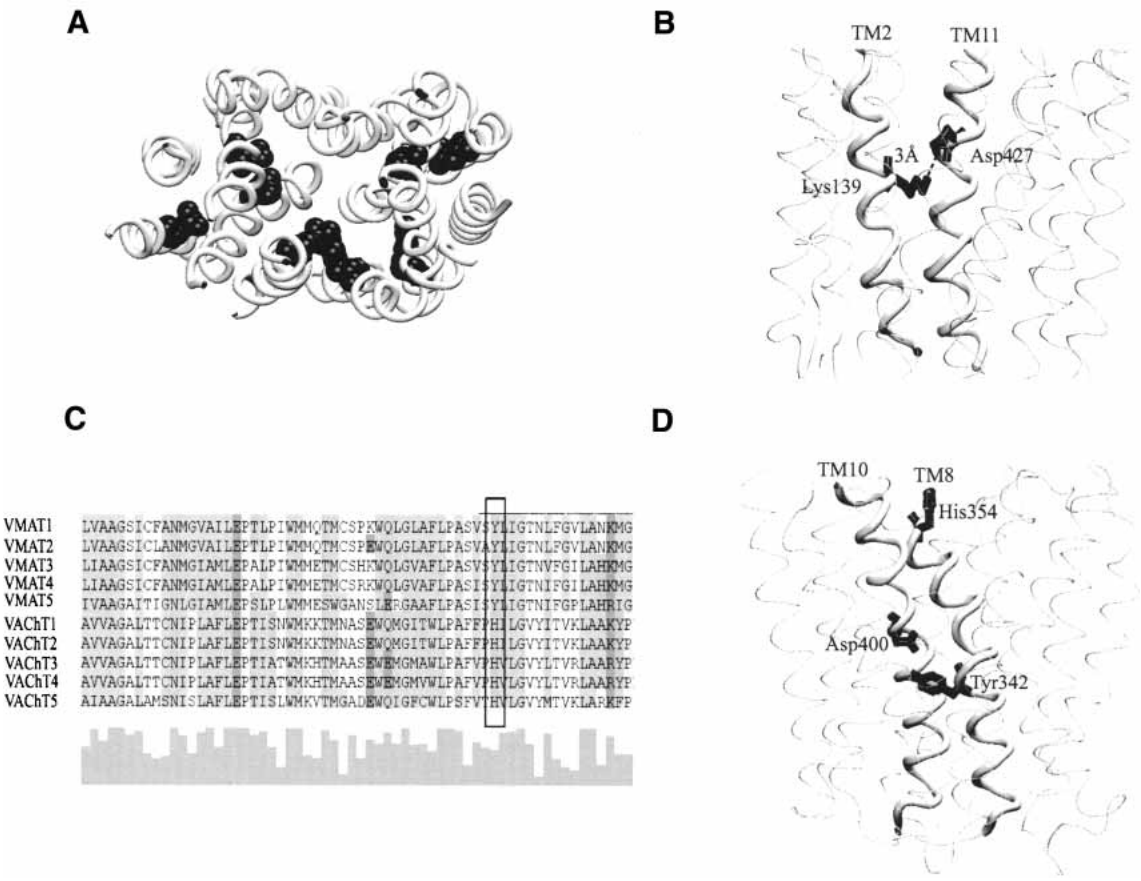

Figure 5. (*A*) Membrane-embedded charged residues in rVMAT2 model. (*B*) Suggested salt bridge between Asp 427 in TM11 and Lys 139 in TM2 of VMAT as it is shown in the model. (*C*) Multiple sequence alignment of VMATs and VAChTs show that Tyr in TM8 of VMATs is replaced by His in VAChTs. (*D*) Shown are TM8 and TM10, Tyr 342 in TM8 (which corresponds to His in VAChT) is close to Asp 400 in TM10.

by GlpT and residues 24 and 253 in the TetAB model by LacY do not face the central cavity of the model, although they are NEM accessible according to biochemical data. Residues that do not agree with NEM accessibility data may suggest that those points are relevant to movement of the protein during the transport cycle or to a local difference in helix bending or stretching. TM7 may serve as an example for such a case because there are more possible NEM-accessible residues, according to the model, than there are in the data (Fig. 4B). The periplasmic half of TM4 and the cytoplasmic half of TM1 are NEM inaccessible, even though they are facing the central cavity of the protein. Those helices in the models are blocked from the cavity by two other helices; TM4 is blocked by TM1 and TM2, whereas TM1 is blocked in a similar way by TM4 and TM5 (Fig. 4). That pattern in the N-terminal lobe continues in the C-terminal lobe of the protein, where TM7 and TM10, which symmetrically correspond to TM1 and TM4, are blocked in the same way by the corresponding TMs (TM7 by TM10–TM11, TM10 by TM7–TM8; Fig. 4).

The charge distribution of the model suggests that the cavity is mostly negative. That feature is in accordance with the positive charge of the GlpT cavity (Huang et al. 2003). The Pi transported by GlpT is negatively charged, whereas Mg^{++}-tetracycline is positively charged. The charge distribution of the GlpT and TetAB model seems to fit their substrate specificity.

The two models are not identical and residues in some places are shifted by one to two helix turns between the two models. The fact that the two models differ from each other indicates that the alignment achieved after manual optimization is not optimal and the exact positioning of the residues is questionable. Still, the accessibility pattern of the helices in both of the suggested models fits the data and the fold is shown to be suitable for this protein. Based on their extensive work on TetAB, Yamaguchi and collaborators (Tamura et al. 2001, 2003) postulated a model for helix packing different from the one shown here. This model proposes a symmetric structure with a central cavity. However, some of the peripheral helices

are located at positions that are different from those suggested here.

rVMAT2 model

The structural data concerning VMAT are not as comprehensive as for TetAB. Because of the lack of experimental data, the structural model of VMAT is less reliable than the model of TetAB. Additional data describing the closely related VAChT were used for further validation. The model was built only to satisfy VMAT data, meaning that all membrane-embedded charged residues are facing the central cavity and the Lys 139 in TM2 is facing Asp 427 in TM11 to enable a salt bridge. The 2-aminoethyl methanethiosulfonate hydrobromide (MTSEA) reaction with Cys 439 in human VMAT2 is inhibited by tetrabenazine (TBZ) binding, suggesting that this residue is close to the TBZ binding site (Thiriot and Ruoho 2001). In the rVMAT2 model, the corresponding residue, Cys 431, is positioned toward the central cavity, one helix turn away from the functionally important Asp 427. Asp 461 in TM12 is related to the protein's apparent affinity to several substances (Peter et al. 1996; Finn and Edwards 1997). In the model, Asp 461 is close to the cytoplasmic side of the transporter and is hidden from the central cavity. Phe 464 in the same helix is related to the apparent affinity of the protein to serotonin (Peter et al. 1996; Finn and Edwards 1997). According to the model, Phe 464 is located in the interface with TM7 in close proximity to Phe 304. The role of those residues in TM12, according to the data, is likely structural and that concept is supported by the model.

VAChT is an MFS protein closely related to VMAT, and biochemical data from VAChT research are often used to support data from VMAT studies (Parsons 2000). Kim et al. (2000) suggested that there is a salt bridge between His 338 in TM8 and Asp 398 in TM10. There is a His in TM8 in the rVMAT model, but it is too far from the Asp in TM10 and cannot interact directly with it. A multiple sequence alignment of VAChT and VMAT revealed that His 338 of VAChT corresponds to Tyr 342 of VMAT. The distance between Tyr 342 and Asp 400 is ~3 Å in the model of VMAT (Fig. 5D), a distance that enables the formation of a salt bridge between two opposite charges.

We demonstrated, with the aid of comparative modeling, that the fold observed for LacY and GlpT also satisfies the experimental and evolutionary data for two other members of the superfamily, VMAT and TetAB. This work supports the line of reasoning presented by Hirai et al. (2003) and the idea that structural conservation of the MFS proteins is high despite the low sequence homology. The simple modeling approach used here may provide a good starting point for research of new MFS proteins. Three-dimensional models, such as those discussed here, may be helpful for biochemical research because they may direct work toward educated guesses. Because the substrate binding site is localized in the cavity between the two lobes, residues facing the cavity are obvious candidates for mutagenesis and biophysical studies. Although the information supplied by the models will simplify the processing and understanding of biochemical data, the limitation and low accuracy of these models must be taken into consideration as well.

Materials and methods

Generating ideal models based on structural alignment

Assessment of comparative models of MFS proteins was done by comparing their known structures and optimal models, which were generated by the use of structural alignment as input. Structural alignments were calculated using the combinatorial extension method (Shindyalov and Bourne 1998), which aligns residues that correspond spatially, without considering sequence similarity. On the basis of those structural alignments, models of the experimentally determined structures of LacY and GlpT were generated, using the respective other protein as template.

Generating an alignment for modeling

To achieve an alignment as close to optimal as possible, we generated a sequence alignment of target and template and manipulated it basically as suggested by Sali, Marti-Renom, and collaborators (Marti-Renom et al. 2000): Twenty randomly chosen MFS sequences of 20% to 40% homology with TetAB or rVMAT2 were aligned with ClustalW (Thompson et al. 1994) against TetAB, rVMAT2, LacY, and GlpT. A desired pairwise alignment of a template protein—LacY or GlpT—with a target protein—TetAB or VMAT—was extracted from the multiple sequence alignment (LacY was also modeled, in order to test the modeling approach). Because of the poor sequence homology, two additional steps were required to optimize the initial alignment for each pair (Fig. 1): (1) overlapping of TM regions predicted for the modeled protein and TM regions of the known structure and (2) fine-tuning of the alignment using experimental data on NEM accessibility or based on the assumption that charged residues in membrane domains should face the hydrophilic cavity. During each modeling cycle, the alignments were modified so that no buried charged residues occur. For TetAB and LacY models, NEM-accessible residues were realigned in a manner that enabled them to face the water-filled cavity of the model.

Prediction of TM regions of MFS proteins was performed with the TMpred (Hofmann and Stoffel 1993) and the TMHMM (Krogh et al. 2001) algorithms. During the optimization process, no gaps were allowed in the predicted or known TM regions.

The long loop between TM1 and TM2 (L1–2; residues 57 to 101 in rVMAT2) in VMATs was not modeled because it is unique to VMAT and VAChT (no homology with other MFSs).

In order to validate the modeling approach, an additional model for LacY was constructed, on the basis of sequence alignment rather than structural alignment, using GlpT as a template. The optimization steps described earlier were used to improve the alignment. The sequence alignment was tuned without taking into consideration the known structure of LacY. The model, produced

from adjusted sequence alignment, was compared with the model based on structural alignment.

Comparative modeling

Comparative modeling was performed with the Modeller 6.2 software (Sali and Blundell 1993). Modeling was performed with the default "model" routine using LacY and GlpT (PDB ID-1PV6 and 4PW4, respectively) as templates and the alignment of each of them with rVMAT2 or TetAB as input. The first five residues and residues 227 to 239 in the L6–7 region of GlpT are not resolved in the structure (Huang et al. 2003) and were therefore omitted from the sequence in the alignments (replaced with gaps).

Analysis of the models

The final alignments were used to generate two models for each of the proteins TetAB and rVMAT2. Similarity of the models and the differences between them were examined by a Cα-RMSD calculation and by visual inspection of the overlapping models, using Swiss pdb viewer and VMD programs (Humphrey et al. 1996; Guex and Peitsch 1997).

The accuracy potential of the MFS models was assessed by comparing optimal models, based on structural alignment, of GlpT and LacY to their solved structures.

To test the ability to reproduce a modified sequence alignment as similar as possible to the structural alignment, we compared the two earlier mentioned LacY models.

Acknowledgments

This work was supported by grants NS16708 from the National Institutes of Health and 463/00 from the Israel Science Foundation. We thank Yael Elbaz for helpful discussions.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H.R., and Iwata, S. 2003. Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* **301:** 610–615.

Ambudkar, S.V., Kimchi-Sarfaty, C., Sauna, Z.E., and Gottesman, M.M. 2003. P-glycoprotein: From genomics to mechanism. *Oncogene* **22:** 7468–7485.

Aronson, H.E., Royer Jr., W.E., and Hendrickson, W.A. 1994. Quantification of tertiary structural conservation despite primary sequence drift in the globin fold. *Protein Sci.* **3:** 1706–1711.

Blakely, R.D. 2001. Physiological genomics of antidepressant targets: Keeping the periphery in mind. *J. Neurosci.* **21:** 8319–8323.

Finn 3rd, J.P. and Edwards, R.H. 1997. Individual residues contribute to multiple differences in ligand recognition between vesicular monoamine transporters 1 and 2. *J. Biol. Chem.* **272:** 16301–16307.

Guex, N. and Peitsch, M.C. 1997. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18:** 2714–2723.

Hirai, T., Heymann, J.A., Shi, D., Sarker, R., Maloney, P.C., and Subramaniam, S. 2002. Three-dimensional structure of a bacterial oxalate transporter. *Nat. Struct. Biol.* **9:** 597–600.

Hirai, T., Heymann, J.A., Maloney, P.C., and Subramaniam, S. 2003. Structural model for 12-helix transporters belonging to the major facilitator superfamily. *J. Bacteriol.* **185:** 1712–1718.

Hofmann, K. and Stoffel, W. 1993. TMBASE—A database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler* **347:** 166.

Huang, Y., Lemieux, M.J., Song, J., Auer, M., and Wang, D.N. 2003. Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*. *Science* **301:** 616–620.

Humphrey, W., Dalke, A., and Schulten, K. 1996. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14:** 33–38, 27–28.

Kaback, H.R., Sahin-Toth, M., and Weinglass, A.B. 2001. The kamikaze approach to membrane transport. *Nat. Rev. Mol. Cell Biol.* **2:** 610–620.

Kim, M.H., Lu, M., Kelly, M., and Hersh, L.B. 2000. Mutational analysis of basic residues in the rat vesicular acetylcholine transporter. Identification of a transmembrane ion pair and evidence that histidine is not involved in proton translocation. *J. Biol. Chem.* **275:** 6175–6180.

Kimura, T., Suzuki, M., Sawai, T., and Yamaguchi, A. 1996. Determination of a transmembrane segment using cysteine-scanning mutants of transposon Tn10-encoded metal-tetracycline/H+ antiporter. *Biochemistry* **35:** 15896–15899.

Kimura, T., Shiina, Y., Sawai, T., and Yamaguchi, A. 1998. Cysteine-scanning mutagenesis around transmembrane segment III of Tn10-encoded metal-tetracycline/H+ antiporter. *J. Biol. Chem.* **273:** 5243–5247.

Kimura-Someya, T., Iwaki, S., Konishi, S., Tamura, N., Kubo, Y., and Yamaguchi, A. 2000. Cysteine-scanning mutagenesis around transmembrane segments 1 and 11 and their flanking loop regions of Tn10-encoded metal-Tetracycline/H+ antiporter. *J. Biol. Chem.* **275:** 18692–18697.

Konishi, S., Iwaki, S., Kimura-Someya, T., and Yamaguchi, A. 1999. Cysteine-scanning mutagenesis around transmembrane segment VI of Tn10-encoded metal-tetracycline/H(+) antiporter. *FEBS Lett.* **461:** 315–318.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305:** 567–580.

Kwaw, I., Zen, K.C., Hu, Y., and Kaback, H.R. 2001. Site-directed sulfhydryl labeling of the lactose permease of *Escherichia coli*: Helices IV and V that contain the major determinants for substrate binding. *Biochemistry* **40:** 10491–10499.

Liu, Y., Peter, D., Roghani, A., Schuldiner, S., Prive, G.G., Eisenberg, D., Brecha, N., and Edwards, R.H. 1992. A cDNA that suppresses MPP+ toxicity encodes a vesicular amine transporter. *Cell* **70:** 539–551.

Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. 2000. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **29:** 291–325.

Merickel, A., Kaback, H.R., and Edwards, R.H. 1997. Charged residues in transmembrane domains II and XI of a vesicular monoamine transporter form a charge pair that promotes high affinity substrate recognition. *J. Biol. Chem.* **272:** 5403–5408.

Parsons, S.M. 2000. Transport mechanisms in acetylcholine and monoamine storage. *FASEB J.* **14:** 2423–2434.

Paulsen, I.T. 2003. Multidrug efflux pumps and resistance: Regulation and evolution. *Curr. Opin. Microbiol.* **6:** 446–451.

Peter, D., Vu, T., and Edwards, R.H. 1996. Chimeric vesicular monoamine transporters identify structural domains that influence substrate affinity and sensitivity to tetrabenazine. *J. Biol. Chem.* **271:** 2979–2986.

Putman, M., van Veen, H.W., and Konings, W.N. 2000. Molecular properties of bacterial multidrug transporters. *Microbiol. Mol. Biol. Rev.* **64:** 672–693.

Rupasinghe, S., Baudry, J., and Schuler, M.A. 2003. Common active site architecture and binding strategy of four phenylpropanoid P450s from *Arabidopsis thaliana* as revealed by molecular modeling. *Protein Eng.* **16:** 721–731.

Saier Jr., M.H. and Paulsen, I.T. 2001. Phylogeny of multidrug transporters. *Semin. Cell Dev. Biol.* **12:** 205–213.

Saier Jr., M.H., Beatty, J.T., Goffeau, A., Harley, K.T., Heijne, W.H., Huang, S.C., Jack, D.L., Jahn, P.S., Lew, K., Liu, J., et al. 1999. The major facilitator superfamily. *J. Mol. Microbiol. Biotechnol.* **1:** 257–279.

Sali, A. and Blundell, T.L. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234:** 779–815.

Schuldiner, S., Shirvan, A., and Linial, M. 1995. Vesicular neurotransmitter transporters: From bacteria to humans. *Physiol. Rev.* **75:** 369–392.

Shindyalov, I.N. and Bourne, P.E. 1998. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* **11:** 739–747.

Tamura, N., Konishi, S., Iwaki, S., Kimura-Someya, T., Nada, S., and Yamaguchi, A. 2001. Complete cysteine-scanning mutagenesis and site-directed chemical modification of the Tn10-encoded metal-tetracycline/H+ antiporter. *J. Biol. Chem.* **276:** 20330–20339.

Tamura, N., Konishi, S., and Yamaguchi, A. 2003. Mechanisms of drug/H+ antiport: Complete cysteine-scanning mutagenesis and the protein engineering approach. *Curr. Opin. Chem. Biol.* **7:** 570–579.

Thiriot, D.S. and Ruoho, A.E. 2001. Mutagenesis and derivatization of human vesicle monoamine transporter 2 (VMAT2) cysteines identifies transporter

domains involved in tetrabenazine binding and substrate transport. *J. Biol. Chem.* **276:** 27304–27315.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673–4680.

Venkatesan, P., Hu, Y., and Kaback, H.R. 2000a. Site-directed sulfhydryl labeling of the lactose permease of *Escherichia coli*: Helix X. *Biochemistry* **39:** 10656–10661.

Venkatesan, P., Kwaw, I., Hu, Y., and Kaback, H.R. 2000b. Site-directed sulfhydryl labeling of the lactose permease of *Escherichia coli*: Helix VII. *Biochemistry* **39:** 10641–10648.

Venkatesan, P., Liu, Z., Hu, Y., and Kaback, H.R. 2000c. Site-directed sulfhydryl labeling of the lactose permease of *Escherichia coli*: N-ethylmaleimide-sensitive face of helix II. *Biochemistry* **39:** 10649–10655.

von Heijne, G. and Gavel, Y. 1988. Topogenic signals in integral membrane proteins. *Eur. J. Biochem.* **174:** 671–678.

Weinglass, A.B. and Kaback, H.R. 2000. The central cytoplasmic loop of the major facilitator superfamily of transport proteins governs efficient membrane insertion. *Proc. Natl. Acad. Sci.* **97:** 8938–8943.